The Effects of Scene Inversion on Change Blindness

DAVID I. SHORE
RAYMOND M. KLEIN
Department of Psychology
Dalhousie University, Nova Scotia, Canada

ABSTRACT. In two experiments, participants searched for a difference between two views of a scene. In Experiment 1, the authors extended the change-blindness findings from previous work by R. A. Rensink, J. K. O’Regan, and J. J. Clark (1997), which used an experimenter-induced global transient, to a less artificial situation in which participants searched for a difference in a pair of photographic images presented simultaneously. To examine the idea that meaning-driven endogenous orienting was responsible for the previously observed advantage for changes in center-of-interest items, the authors inverted half of the image pairs. The advantage for center-of-interest items was replicated with upright displays, but it was completely eliminated by inversion, strongly supporting the role of meaning-driven endogenous orienting in this task. With flickering displays (Experiment 2), the center-of-interest effect was completely unaffected by inversion. The authors suggest that when change blindness is induced via flicker, scene modifications are typically found by stimulus-driven rather than by meaning-driven processes.

OUR INTERNAL REPRESENTATION of the external world appears to be rich, stable, and complete. Yet, as we move around our environment, or alternatively, as objects in the environment move around us, the image reaching the visual cortex changes drastically. Despite this constant change, we perceive a coherent and steady world that seems to contain all of the detail present in the environment. This comprehensive representation needs to be considered in the context of the need to select and process only a subset of the available visual information (cf. Enns, 1992). There is an infinite amount of information in the visual world at any one instant, and we can process only a finite amount of it. Some information must remain relatively unprocessed. These two facts, our rich subjective experience and the need for selectivity, present something of a contradiction. How can we represent the entire visual scene and yet process only some of it? This contradiction between what we perceive and what must be true about selectivity seems to imply that our subjective impression of a comprehensive representation is an illusion (cf. O’Regan, 1992; Stroud, 1955).
The richness of our phenomenological experience and the notion that this may be an illusion have been discussed in the literature on both memory and perception for well over a century (e.g., Helmholtz, 1867/1962). A different emphasis was espoused by Bartlett (1932), who proposed that most of the richness in our representation was supplied by memories of previous environments similar to the present one; that is, we fill in, from past experience, aspects of the world of which we are not immediately aware.

More recently, empirical work has crystallized this issue by asking what information can be integrated from one fixation to the next. We explore our visual world by fixating on numerous locations in sequence. Any spatiotopic representation (i.e., one coded in environmental rather than retinotopic coordinates) must begin by integrating information from two fixations that are separated by a saccade. Such integration has been termed transsaccadic memory (Breitmeyer, Kropfl, & Julesz, 1982; Jonides, Irwin, & Yantis, 1982). There are a number of experiments in the literature claiming to have found evidence for such a memory mechanism, but on closer inspection and after attempts to replicate, it seems that there is no such form of memory (Bridgeman, Hendry, & Stark, 1975; Henderson, 1997; see Irwin, 1991, 1992, 1996, for reviews of this literature). This failure to find transsaccadic memory seems to imply that information about the visual environment is recalculated, perhaps from scratch, upon each fixation. This sentiment was forcefully forwarded by O'Regan (1992), who proposed that there is no need to maintain an accurate memory of the external visual world because that world is always available for inspection. One conclusion from this literature is that our visual system neither has nor needs a spatiotopic memory.

A number of observations consistent with this view show that many changes can go unnoticed if they are made during a saccade. Grimes (1996), for example, modified a naturalistic scene while a participant was making an eye movement. Somewhat surprisingly, very large changes in object color, location, or identity went unnoticed unless that object was the target for the next saccade. That is, unless an object had been specifically targeted, and presumably attended, a change in that object was not seen. Although this was a counterintuitive observation at the time, in light of the failure to find transsaccadic memory and the notion that we do not maintain an accurate spatiotopic representation, this should not have been so surprising.

Following these observations, it has been shown that similar changes can go unnoticed for a number of reasons in addition to the execution of a saccade (see

This research was supported by the Killam Foundation with a postdoctoral fellowship to the first author and by the National Research and Engineering Council of Canada with an operating grant to the second author. The authors thank the students of 3130 for participating in Experiment 1 and for lively discussions concerning the data presented here.

Address correspondence to Raymond M. Klein, Department of Psychology, Dalhousie University, Halifax, Nova Scotia B3H 1J4, Canada; Ray.Klein@dal.ca or dshore@or.psychoogy.dal.ca (e-mail).
Simons & Levin, 1997, for a review; see also special issues of Visual Cognition, Jan.–March 2000, Vol. 7, Nos. 1, 2, 3. Changes will go undetected if they are made across a scene change in a movie (Simons, 1996), presented during a blink (O’Regan, Deubel, Clark, & Rensink, 1997), shown coincident with a number of mud splats (O’Regan, Rensink, & Clark, 1999), or accompanied by a large transient that fills an entire computer screen (Rensink et al., 1997, 2000).

Rensink, O’Regan, and Clark (1997) used pictures of naturalistic scenes and presented these in alternation with the identical scenes, save for a carefully generated modification. In this article, we spend some time describing how these stimuli were generated and what was found with them, because we used these stimuli in our studies. The modification could be of three types: presence/absence of an object, location of an object, or change in the color of an object. Additionally, the changed object could be the center of interest, or of marginal interest in the scene. Five naive participants were asked to describe each picture, and any item that was mentioned by more than three participants was considered a center of interest; all other items were considered of marginal interest (Rensink et al., 1997). In the critical condition, the two images were presented repeatedly in sequence with an intervening white, blank screen. As long as the blank screen was presented for a sufficient duration (around 80 ms), the change was very difficult to detect. The time to detect the change was modulated by both manipulated factors—type of change and center of interest (see Figure 1). Items of marginal interest took longer to detect, regardless of the type of change. The relative difficulty in detecting the change in this “flicker” condition was compared with a control condition, which did not contain the intervening blank screen. Without the global transient associated with the blank screen, performance was uniformly fast across all conditions (see Figure 1). That is, without the 80 ms of blank screen, the object change was immediately evident because it produced a unique transient that attracted attention. With the intervening blank screen, the change could not be detected because the representation of the first image was overwritten or substituted by the second (cf. Enns & Di Lollo, 1997). Rensink et al. (1997) postulated that the masking effects of the global transient could be avoided only if attention was directed to the relevant location in the scene prior to its removal, allowing the smaller object-based transient to become evident. The effect of interest (faster detection of central-interest than of marginal-interest objects) was explained by the allocation of attention in a volitional fashion using the meaning of the scene to direct scanning (Rensink et al., 1997, p. 372).

The first goal of the present research was to test this claim using a manipulation—inversion—known to disrupt the extraction of meaning. Rock (1974) has argued cogently that changes in orientation in general, and inversion in particular, dramatically alter the meaning of objects and scenes. To support this argu-

---

1We are grateful to Ronald A. Rensink for supplying the images used in the present set of experiments.
FIGURE 1. Reaction times from the original study (Rensink et al., 1997), showing the effect of interest and type of change. Also shown is the control condition, which had no intervening blank screen between the two versions of the scenes. Note. From “To See or Not to See: The Need for Attention to Perceive Changes in Scenes,” by R. A. Rensink, J. K. O’Regan, and J. J. Clark, 1997, *Psychological Science, 8*, p. 371. Copyright 1997 by Blackwell Publishers. Reprinted with permission.

ment he presented a number of convincing demonstrations that inverted objects are often not recognized as the familiar things that they are (e.g., a map of Africa or a photo of a famous president). Beyond these phenomenological demonstrations are two sources of empirical support. Intraub (1984) reasoned that inverted images should produce less conceptual masking—reduced recognition memory of a target photo that was immediately followed by an unrelated photo—than similar photos presented upright. She found a large (but nonsignificant, *p* < .10) reduction in conceptual masking when the mask was presented upside-down. Finally, Klein (1982) has shown that ratings of configural similarity for pairs of briefly presented figures, identical except for 180° rotation, are most detrimentally affected by inversion, compared with reflections in the other two axes. To the extent that inversion disrupts the extraction of meaning and that the advantage for center-of-interest items in the Rensink et al. (1997) study was caused by meaning-driven control of attention, scene inversion should reduce this advantage, possibly to the point of producing a similar pattern of change blindness for both the central and marginal interest objects.
The flicker paradigm evolved from the research on saccade-contingent change and was developed to allow for more normal viewing conditions (Rensink et al., 1997, p. 368), since no eye-tracking equipment or chin rest was necessary to evidence change blindness. The second goal of the present research was to make the task even more similar to normal viewing by allowing participants full control over the scanning demands and timing. This was accomplished by presenting the two images simultaneously on a laminated piece of paper and having participants detect differences as opposed to changes. There are several advantages to this technique over the flicker paradigm. First, participants were under truly “normal viewing conditions,” because there was no eye-tracking equipment, no chin rest, and even more natural, the computer-generated image modifications were not presented online. The participant simply sat at a desk examining a piece of paper with two images on it, searching for a difference. The second advantage has to do with the need for a transient in the flicker paradigm and its proposed role in simulating an eye movement. In the present experiment, the only action that may have disrupted any representations of the scene that participants may have formed would have been saccadic eye movements. The simultaneous presentation of two slightly different images to illustrate the paucity of our visual representations during scene inspection has a long history in comics and cereal boxes and has been discussed in the literature (Irwin, 1991, 1992; see also the cover of Trends in Cognitive Science, Vol. 1, No. 7, where this form of display was used to showcase an article on this topic by Simons and Levin, 1997). Although this method has also been explored empirically (Brunel & Ninio, 1997; Orbach & Scott-Brown, 1998), these experiments have typically used artificial stimuli and forced participants to maintain a steady fixation. As far as we know, ours is the first systematic, empirical investigation using such naturalistic displays and making direct comparisons between this method and the flicker technique.

EXPERIMENT 1

Method

Participants

Fifteen participants (10 women, 5 men) were recruited by word of mouth from the population of research assistants, graduate students, and undergraduate students in the Department of Psychology at Dalhousie University. Some were paid $5 for their participation, others received course credit in Psychology 1000, and the remaining volunteered their time for no remuneration. All were naive to the purpose of the experiment, and none had seen the stimuli previously. All had normal or corrected-to-normal vision. One participant was excluded from the analysis because he reported that he was color blind.
Forty-eight pictures, which were used by Rensink et al. (1997), were reduced to 5" × 7" (12.70 cm × 17.78 cm) and printed on 8.5" × 11" (21.59 cm × 27.94 cm) photo quality Epson paper with an Epson Color 800 printer. Every image had two versions: the original and a modified copy. What was modified and how it was modified were determined by crossing Interest (central or marginal) and Type of Change (color, location, or presence/absence). Both versions (modified and unmodified) of each scene were printed on the same page, with the original placed randomly on the top or bottom half of the page. These pages were then laminated using a pouch lamination procedure. Half of these pages were presented upright, and half were presented inverted.

The picture pairs were divided into four sets of 12 pictures. Each set contained 1 exemplar of the conditions represented by the factorial combination of type of change, interest, and orientation. The 12 pictures in a set were randomly ordered at the beginning of the experiment, and the same order or reverse was used for each participant. After each participant finished with the entire set, those pictures were inverted so that the next participant received the same order but the opposite orientation of each image. After 7 participants completed the experiment, the order of all 48 images was reversed.

Procedure

The participant first read an instruction sheet and signed an informed consent form. Then, we reviewed the instructions with the participant and answered any questions. We discussed each type of change and highlighted the fact that half of the pictures would be inverted. We began each trial by saying, “Ready . . . go,” and starting a hand-held stopwatch. The participants turned over the first sheet and searched for the difference between the two pictures. Once the difference was found, they pointed it out and gave a brief verbal description of the change. Once the participant indicated the difference, the stopwatch was stopped. On the few trials that participants were incorrect (<1% of all trials), we told them that this was not the difference and continued the timing until either they detected the correct difference or 1 min had elapsed. If, after 1 min, the participant had not found the difference, the stopwatch was stopped, and we indicated the difference. A time of 60 s was assigned for that trial. Upon completion of the trial, we recorded the time and began the next trial. This continued until all 48 images had been searched.

Results and Discussion

Participants failed to detect the difference in 9.6% of trials. The mean reaction times from all 12 conditions are shown in Figure 2. Figure 2A presents the data from the upright condition, and the data from the inverted condition are shown in Figure 2B. These data were submitted to a three-factor repeated measures analysis of variance (ANOVA): Interest (central, marginal) × Type of Change
Figure 2. Mean reaction times from Experiment 1 for upright (A) and inverted (B) versions of the scenes that were presented simultaneously on a single sheet of paper. Data are presented for conditions represented by the factorial combination of type of change and interest. Error bars indicate the between-subjects standard error of the mean.

(presence/absence, color, location) × Orientation (upright, inverted). The type of change had a clear influence on the reaction time, $F(2, 26) = 27.7$, $MSE = 106.0$, $p < .0001$, with location changes taking longer than both color and presence/absence changes, which did not differ from each other. There was also a main effect of interest, $F(1, 13) = 9.7$, $MSE = 38.4$, $p < .01$, which indicated that, overall, changes to central-interest objects were detected faster than changes to marginal-interest objects. This main effect was moderated by two higher order interactions: type of change, $F(2, 26) = 4.8$, $MSE = 49.2$, $p < .025$, and orientation of the images, $F(1, 13) = 13.3$, $MSE = 30.8$, $p < .005$. The upright images showed a robust effect of interest, $F(1, 13) = 26.5$, $MSE = 29.5$, $p < .0005$, whereas the inverted images showed no effect of interest, $F(1, 13) = .01$, $MSE = 39.7$, ns.
This experiment replicated the findings from the previous study (Rensink et al., 1997) using these pictures for upright images and, by extending the task to one that did not require any high-technology equipment, the experiment demonstrated that the change blindness phenomenon is accessible in a real-world task using an extremely simple methodology. The task also highlights the observation made earlier that an eye movement alone is sufficient to disrupt any complete representation of the scene (Irwin, 1992). The similar pattern of results across type of change observed here (see also Experiment 2) and in the Rensink et al. (1997) study seems to indicate that making changes over space or time has similar consequences in terms of both the time to find a difference and the effect of different types of change. Of course, in the present experiment, in which space was the manipulated variable, time was inherently necessary to make eye and attentional movements. Whereas in all of the previous research there was some transient aspect to the procedure that either helped or hindered participants’ performance in change detection, one unique aspect of the present research is that the only transient in it was produced by the participant (i.e., there was no experimenter-induced transient). We return to this point in the General Discussion.

The theoretically more important finding is that inverting the images eliminated the influence of interest. That inversion reduces the efficacy with which meaning can be extracted from an image is consistent with the conclusion reached by Rensink et al. (1997) that meaning-driven endogenous control is responsible for the effect of interest. The basic logic follows that objects that are important to the meaning of the image are scanned first, and thus a change in one of them is detected before a change in a component of the scene that is unimportant to its meaning.

As noted earlier, the major difference between the present procedure and that in the flicker paradigm is the involvement of an experimenter-induced transient in the latter method. In order to be certain that the role of meaning found with simultaneous presentation is also found in the flicker paradigm, we replicated that procedure with the manipulation of inversion.

EXPERIMENT 2

This experiment replicated Experiment 2 of Rensink et al. (1997) with very similar timing parameters. The manipulated factors and their combinations (type of change and orientation) were identical to those used in Experiment 1. To reiterate, the key predictions concerned the interaction between orientation and interest. To the extent that meaning-driven endogenous orienting is the cause of the interest effect (faster detection for central-interest than marginal-interest objects) and that inversion disrupts meaning, inverted images should show a reduced effect of interest, as was found in Experiment 1.
Method

Participants

Eighteen students in a 3rd-year psychology course at Dalhousie University participated as part of a course laboratory exercise. Participation in the experiment was compulsory, but students were given the option of having their data excluded from the final data set. This informed consent was collected in such a manner that neither the teachers of the course nor the researchers were aware of the students’ wishes. All 18 students (2 men and 16 women) agreed to have their data included. All had normal or corrected-to-normal vision, and none was color blind.

Apparatus and Stimuli

Pictures were presented on an Apple 15” RGB monitor connected to a Macintosh Power PC running VScope software. Viewing distance was about 50 cm, and participants were free to move their heads. At that distance, the screen subtended 27° × 18° of visual angle. Each picture occupied the entire screen. The 48 images used were the same set of pictures used in Experiment 1 and by Rensink et al. (1997, 2000).

Procedure

All participants performed the task simultaneously in a single room. Trials were self-initiated and self-terminated. When the participant pressed the space bar, the screen went blank for 60 ms; then the original version of the image was presented for 555 ms, followed by a blank screen for 120 ms. The modified version was then presented for 555 ms and was followed by a blank screen for 120 ms. The two images alternated in this fashion until the participant responded by pressing the space bar or until 1 min had passed, after which time the trial was terminated, and 60 s was entered as the response time. We instructed the participants to press the space bar as soon as they had detected the object that had been changing in the image. After we gave these instructions, we answered any questions.

The 48 images were divided into four blocks of 12 images, which were displayed in random order. Each block contained 1 example of the 12 conditions, Orientation (2) × Interest (2) × Type of Change (3), which meant that each condition was presented before any condition could be presented again. The two groups were identical in all respects except for which group of images was inverted.

Results and Discussion

Participants failed to detect the difference in 1.6% of trials. The mean detection times for all 12 conditions are shown in Figure 3. These reaction times were
entered into a three-factor repeated measures ANOVA: Interest (central, marginal) × Type of Change (presence/absence, color, location) × Orientation (upright, inverted). The graph reveals that, overall, changes in objects of central interest were detected faster than changes in objects of marginal interest, $F(1, 17) = 83.0, MSE = 20.1, p < .0001$. Detection of object presence/absence occurred faster than detection of color change, which in turn occurred faster than detection of change in location, $F(2, 34) = 3.9, MSE = 16.5, p < .05$. These two factors (type of change and center of interest) also entered into a significant interaction, $F(2, 34) = 28.7, MSE = 16.6, p < .0001$, indicating that the effect of central versus marginal interest was modulated by the type of change. Specifically, the difference between central and marginal interest was greatest for location change and smallest for color.

**FIGURE 3.** Mean reaction times from Experiment 2 for upright (A) and inverted (B) versions of the scenes that were presented in the flicker paradigm on a computer screen. Data are presented for conditions represented by the factorial combination of interest and type of change. Error bars indicate the between-subjects standard error of the mean.
change, with presence/absence falling between these two. This pattern closely replicates the results found by Rensink et al. (1997, Figure 3A; see Figure 1).

Most important, the patterns of data for the two orientations were extremely similar in both form and magnitude. There was no main effect of orientation, \( F(1, 17) = .32, MSE = 4.9, ns \), and orientation did not enter into any interactions, \( F(2, 34) = .03, MSE = 19.96, ns \), for the three-way interaction between type of object, type of change, and orientation. Because of the striking difference between these results and those of Experiment 1 with respect to the interaction of orientation and interest, we conducted a between-experiment analysis.

**Comparing Results of Experiment 1 and Experiment 2**

A four-factor mixed design was used to examine the patterns of data across the two experiments: Interest (central, marginal) \( \times \) Type of Change (presence/absence, color, location) \( \times \) Orientation (upright, inverted) \( \times \) Experiment (flicker, paper). The first three factors were repeated measures, whereas the last was between-subjects. This analysis confirmed those performed on the individual experiments, indicating that interest, \( F(1, 30) = 60.2, MSE = 27.4, p < .0001 \), and type of change, \( F(2, 60) = 33.8, MSE = 55.9, p < .0001 \), had robust effects on the time to detect differences and that these two factors entered into a significant interaction, \( F(2, 60) = 20.5, MSE = 31.0, p < .0001 \), reflecting that the effect of interest, while consistent, was different for the three types of change. This combined analysis revealed a main effect of experiment, \( F(1, 30) = 45.0, MSE = 230.7, p < .0001 \), with Experiment 1 evidencing reaction times almost twice those found in Experiment 2. Note the different scales for Figures 2 and 3. The most important result from this analysis is the observation that the effect of interest was moderated by orientation in Experiment 1 but not in Experiment 2, \( F(1, 30) = 7.3, MSE = 27.2, p < .025 \), for the three-way interaction of experiment, orientation, and interest. This three-way interaction is exemplified in Figure 4, in which the data are collapsed across the type of change.

Rensink et al. (1997) attributed the detection of change in the flicker paradigm to volitional control (p. 372). In contrast, our results are consistent with the proposal that meaning-driven endogenous control of scanning was not responsible for the difference between central- and marginal-interest changes found in the flicker paradigm. This conclusion depends critically on our assertion that inversion disrupts the encoding of meaning from the scenes we presented. Despite the evidence reviewed earlier, we thought it prudent to demonstrate this effect directly using the same stimuli that generated the data shown in Figures 2, 3, and 4. Therefore, we conducted a simple empirical test of the assumption that the extraction of meaning either is more difficult or takes longer when the images are inverted. We asked 5 participants to describe the contents of the scenes when they were presented upright or inverted. Additionally, the scenes were either flickered at the same rate used in Experiment 2 or presented continuously for up to 4 s. The par-
Participants initiated and terminated the presentation of the image, and we instructed them that they should limit their viewing time to as short a time as possible, while still providing an adequate description. All 5 participants chose to view the inverted pictures for longer than the upright pictures, $F(1, 4) = 7.9$, $MSE = 32927.6$, $p < .05$. There was no effect of presentation mode (flicker vs. continuous viewing), nor did this factor interact with inversion, $Fs < 1.0$. We also examined the descriptions given by the participants to ensure that the center-of-interest definitions used for the upright versions of the pictures applied to the inverted images. Using the same criterion as Rensink et al. (1997), we observed no difference in the definition of interest. This supplemental study revealed two important results: Extraction of the meaning of the scenes is adversely affected by inversion, and this disruptive effect applies equally to continuously presented and flickering scenes.

**GENERAL DISCUSSION**

One rationale for conducting these experiments was to determine the degree to which meaning-driven endogenous control is responsible for the detection of experimenter-induced changes. Inversion was proposed as a manipulation that should interfere with the extraction and use of meaning while maintaining stimulus salience. The finding in Experiment 1 that the center-of-interest advantage was
eliminated by scene inversion suggests that when an individual examines images of a scene presented simultaneously (side by side), the inspection process is guided by meaning. In contrast, the lack of an inversion effect in Experiment 2, in which the two images were presented in the flicker paradigm (in alternation, in the same location, with intervening blanks), seems to imply that meaning does not play a particularly important role in the flicker paradigm. Before we discuss this proposal at length, it behooves us to explain how a center-of-interest effect could be observed when meaning may not be playing a role.

In this regard, it is important to note that in the stimulus set developed by Rensink et al. (1997) and used here, the meaningfulness of a particular object in the scene and its salience cannot be confidently separated. The central-interest objects could have been described as such because they were important to the meaning of the scene, or because they were physically salient, or for both reasons. The rating procedure used by participants in the original study (Rensink et al., 1997) does not allow an assessment of the relative contributions of these two factors. Thus it is possible that the center-of-interest effect observed with flicker (and unaffected by inversion) was attributable to the physical salience of the components labeled as center of interest as opposed to their meaning within the scene.

**Endogenous and Exogenous Factors in Scene Scanning**

Visual exploration of a scene involves the intricate interplay between endogenous and exogenous control of visual orienting (Klein, Kingstone, & Pontefract, 1992; Klein & Shore, in press). Stimulus salience affords potential locations in the scene for eye fixations, whereas goal-directed factors influence which of these will be selected. In a study of this interaction, Yarbus (1967) showed that the scanning path of a picture is affected by the question posed to participants. When asked about the age of the people in the “Unexpected Visitor,” participants’ fixations tended toward the faces of the people in the picture, whereas when asked about their relative wealth, participants’ fixations tended to be allocated to the many different objects in the room. Precisely the same stimulus features were available in each context, but the endogenous selection of the features was guided by the question posed about the scene.

We believe that there are two routes by which the natural use of meaning to guide endogenous control of visual orienting may be disrupted or discouraged in the flicker paradigm. One of these is related to motor execution; the other, to information pickup. In normal viewing, the participant controls when transients caused by eye movements occur, and thus the perceptual-motor system can anticipate and account for these transients. Although this occurrence is, to a degree, true in the flicker paradigm (every saccade is associated with a predictable transient), the frequent transients that are not controlled by the oculomotor system may disrupt endogenous control over this system, encouraging a reliance on exogenous control. In contrast, in the simultaneous inspection paradigm, as in
normal viewing, the saccade-induced transients occur during a preplanned voluntary eye movement and hence should not interfere with meaning-driven endogenous control (indeed, they are an expected part of it).

As for the information pickup side of the equation, we suggest that the concept of attentional-control settings (Folk, Remington, & Johnston, 1992) can also help to explain the dissociation we have observed here. In the flicker paradigm, because the two images of the scene are presented successively in the same location, the target is a local change in a particular spatial location. Although this change is, to a degree, overshadowed by the global transient associated with the transitions between the scene and the blank screen, we suggest that the participant adopts an attentional control setting aimed at detecting the local signals associated with the changing scene component. This control setting emphasizes local, low-level changes between successive exposures, while it correspondingly decreases the reliance on meaning to guide orienting. As the participant is looking for a target defined by local, low-level perceptual salience, visual orienting becomes dominated by the salience-driven exogenous control system. In contrast, in the side-by-side, simultaneous paradigm, the difference between the images is not detectable as a local change. Hence, the search is accomplished via encoding into memory the scene components that are attended as a result of meaning-driven endogenous orienting.

Thus, for two complementary reasons, the two methods for eliciting and measuring change blindness force the participants into different modes of processing. In the flicker paradigm, because the target is a local change, an attentional control setting is adopted that gives priority to exogenous control by salient scene components. Concurrently, those non-self-generated transients that are associated with flicker discourage endogenous control of orienting. In the simultaneous paradigm, because transients are self-generated and can be anticipated, there is no disruption of the endogenous control system. Moreover, because the target is a difference rather than a change, the scene components relevant to meaning are the initial targets of attention, and if these are not changing, search among the less central objects ensues. With inversion, meaning is not extracted or is more difficult to extract, and hence interest does not play much of a role.

Speculative confirmation of the relative roles of exogenous and endogenous factors may be found in the raw reaction times across the two experiments. This is speculative because of the many differences between the two procedures other than the absence or presence of a global transient. Nevertheless, in Experiment 2, for which we are proposing that change is detected by exogenous factors attracted to stimulus salience, reaction times were, on average, half that found in Experiment 1, for which we propose that endogenous orienting controls scanning. With endogenous control, scanning must proceed in a systematic fashion from object to object, whereas under exogenous control, search proceeds from one salient feature to the next until the change is found (cf. Theeuwes, Atchley, & Kramer, in press; Wolfe, Cave, & Franzel, 1989). Converging evidence for this argument may
be found in a recent study by Scholl (2000), showing that an exogenous cue (i.e., a color singleton or a late onset), which is just as likely to be presented near a changing item as near a nonchanging item, can attenuate change blindness.

There is an alternative explanation of our data that builds on the fact that we have observed a center-of-interest advantage in all conditions except the simultaneous condition with inversion. Rensink (personal communication, November 1998) notes that unlike the flicker paradigm, which involves change detection, the simultaneous paradigm depends on detecting a difference in objects or other scene components. He further proposes that with the inverted scenes, each time a new object file is set up, the rotation of the old one is lost, so that object interpretations must be rerotated constantly. If this is too much work, the comparison reverts to image-based properties that would not be dependent on meaning, and thus the center-of-interest advantage is eliminated. This explanation, which does not depend on the ad hoc suggestion that meaning and salience may both underlie center-of-interest ratings for the stimuli we used, can account for the elimination of the center-of-interest advantage in the simultaneous paradigm. However, if the center-of-interest advantage is based on meaning-driven endogenous orienting within the flickering display, then a manipulation that disrupts meaning should reduce or eliminate this advantage. Thus, our finding that inversion had no effect on the center-of-interest advantage with flicker, together with our supplementary experiment showing that the extraction of meaning was disrupted with both flickering and steady displays, poses a problem regarding this explanation.

Our explanation of the data must remain speculative because the present stimuli were not designed to separate stimulus salience and goal-directed meaning. Future researchers should explicitly control these elements. If our proposal is correct, the sequence of image components inspected and revealed by the kind of oculomotor scan path collected in Yarbus’s studies should be different in the flicker and simultaneous free-viewing conditions. Specifically, with flicker, eye movements should proceed from areas of highest salience to areas of lower salience, whereas with simultaneous viewing, scanning should proceed from objects key to the meaning of the scene to objects less important to it.

What Is New About Change Blindness?

The main conclusions that we draw from several years of change blindness research are that (a) contrary to what our introspective awareness might suggest, we do not maintain a spatiotopic representation of all of the details in a scene; and (b) it is primarily those features and objects to which we attend (whether this selective process is controlled exogenously or endogenously) that are well enough encoded to be available for comparison with a second image. Put another way, “we only ‘see’ what we attend to” (O’Regan, 1992, p. 473; see Shapiro, 2000, for a similar conclusion). The findings reported here do not undermine this generalization; quite to the contrary, they support and extend it.
This view of perceptual encoding follows naturally from, among other things, the research indicating that there is no transsaccadic memory (see Irwin, 1996). If detailed information cannot be integrated or remembered across saccades, how can it be expected that participants will maintain the infinite amount of detail present in a naturalistic scene? Similar conclusions have been reached during the 40 years since Broadbent (1958) reintroduced attention to the study of behavior. In that seminal work, he also concluded that “only information which passes the filter can be stored for long periods” (p. 242), implying that what is not attended is lost. Around the same time, Sperling’s (1960) studies of iconic memory demonstrated that for items to be available for later report, they must be selected for further processing. Thus, the more recent literature on change blindness (see Simons & Levin, 1997, for a review) can be thought of as expanding and generalizing a historical set of findings from years of rigorous psychophysical research to the inspection of naturalistic scenes. This is an exciting and necessary development if we are to understand how we perceive in everyday life.

REFERENCES


Manuscript received October 27, 1998
Revision accepted for publication January 21, 1999